Energy Conversion and Management 49 (2008) 2142–2151

Contents lists available at ScienceDirect

Energy Conversion and Management

journal homepage: www.elsevier.com/locate/enconman

Comprehensive empirical analysis of ERA-40 surface wind speed distribution over Europe ☆

Péter Kiss *, Imre M. Jánosi

Department of Physics of Complex Systems, Loránd Eötvös University, Pázmány Péter sétány 1/A, Budapest, Hungary

A R T I C L E   I N F O

Article history:
Received 19 June 2007
Accepted 20 February 2008
Available online 10 April 2008

Keywords:
Wind speed
Wind energy
Weibull distribution
Generalized gamma distribution
ERA-40 re-analysis

A B S T R A C T

As a first step of an exhaustive assessment of wind energy potential over Europe, here, we provide a unified description of the wind speed probability distribution both over sea and land. We evaluated surface wind velocity records of the ERA-40 data base covering 44 years with a temporal resolution of 6 hours. We tested the well known distribution functions (Rayleigh, binormal, Weibull, lognormal etc.) and observed that the popular Weibull function performs supremely, however, it fails at many locations over land. We found that the generalized gamma distribution, which has independent shape parameters for both tails, provides an adequate and unified description almost everywhere. The geographical distribution of the fitted parameters reveals the possible climatological origin of different wind speed distributions.

1. Introduction

Wind energy is the world's fastest growing renewable source of electricity; the global capacity has more than quadrupled between 2000 and 2006 [1]. The atmospheric flows are strongly volatile, and therefore, the average wind speed at a given location is a very poor predictor of the energy output of a wind turbine. The basic requirement for wind power estimates is an adequate characterization of the empirical probability distribution of wind speeds since wind direction is less important because of the well developed methods of yaw control for modern turbines [2]. The statistical description is highly simplified when a measured histogram can be accurately fitted by an analytical probability density function (PDF) with a few parameters.

The traditional approach of modeling the wind speed PDF is based on the Rayleigh and the more flexible Weibull distributions [2–8]. However, several authors noted that Weibull fits of empirical data have low quality at several locations, mostly over land [10]. Various analytical forms of skewed distributions were proposed as possible alternatives, such as the lognormal [11,12], square root normal [13,14], chi [11], inverse Gaussian [15], generalized gamma [11], generalized extreme value [16] or extended exponential functions [17,18].

In this work, we report on a detailed analysis of surface wind speed distribution over Europe. The main goal was to find an effective and optimal description of the PDF both for terrestrial and sea surface locations. We evaluated wind data from the ERA-40 re-analysis [19] of the European Centre for Medium Range Weather Forecasts (ECMWF). Our tests unambiguously demonstrate that the generalized gamma (GG) distribution provides an improved fit for an overall statistical characterization of surface wind speed. We show that the GG fit outperforms the Weibull function, especially at the large speed tails, which is the most relevant region with respect to wind power estimates.

After the description of the data and the methods used, we summarise tests with various distributions by fitting empirical wind speed histograms. Besides the fundamental Rayleigh, the commonly used Weibull and lognormal distributions, we show the results for the generalized gamma distribution and demonstrate its superiority. Next, we briefly discuss the spectral properties of wind speed records and analyse wind speed fluctuations by removing seasonal periodicities from the original time series. A summary and conclusions are given in the final section.

2. The data and methods

We evaluated the ECMWF's ERA-40 re-analysis data [19] consisting of the u (eastward) and v (northward) orthogonal components of the horizontal wind field at 10 m above ground level. The data base covers a time period of 44 whole years between 1 September 1958 and 31 August 2002. Four instantaneous values are recorded each day for the main synoptic hours of 00, 06, 12

☆ Wind speed Wind energy Weibull distribution Generalized gamma distribution ERA-40 re-analysis.

* Corresponding author. Tel: +36 1 372 2878; fax: +36 1 372 2866.
E-mail addresses: kisspeter@complex.elte.hu (P. Kiss), janosi@lecso.elte.hu (I.M. Jánosi).

0196-8904/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.enconman.2008.02.003

and 18 UTC at each location. The spatial resolution is $1^\circ \times 1^\circ$ (lat/long), and a given value for an atmospheric variable is considered to be representative for the whole cell. Our analysis is restricted to a geographical area covering Europe: 2501 grid points between 35°N and 75°N latitude and 20°W and 40°E longitude.

The main quantity of interest in this study is the scalar wind speed $s = \sqrt{u^2 + v^2}$. For each grid point, the standard statistical characteristics (mean, standard deviation, skewness and kurtosis) are determined, and then the Rayleigh, binormal, Weibull, lognormal and generalized gamma (GG) distributions are fitted by the standard method of maximum likelihood estimates [5,20]. No preprocessing (filtering or removal of seasonal periodicities) was implemented prior to the computations above. The goodness of fit is characterized by computing the coefficient of determination R^2 (the fraction of the total squared error that is explained by the model). The different models are evaluated by comparing the unexplained percentage variance $100(1 - R^2)$ for a given data set.

In order to characterize the temporal behavior of wind speed records, frequency domain analysis is performed by the usual Fourier spectral method. The effects of seasonal periodicities are evaluated by removing the long range average values for a given hour of a given calendar day from the original time series, as usual.

The average value $\bar{s}$ and standard deviation σ_s of ERA-40 surface wind speeds computed over the whole period of 44 years (Fig. 1) illustrate the gross features of wind climatology over Europe. The strong coupling between the values of average speed and standard deviation is apparent, the coefficient of variation is around $\sigma_s/\bar{s} \approx 0.5$, except for a few isolated regions (for example around Corsica).

3. Models for wind speed histograms

3.1. Rayleigh distribution

The most transparent model for scalar wind speed distribution is based on the assumptions that the orthogonal u and v components are independent and identically distributed (iid) Gaussian random variables with zero means and equal standard deviations of $s_0/\sqrt{2}$ (we adopt this notation to get simpler mathematical formulae below). Of course, all the higher moments (skewness, kurtosis etc.) are identically zero. In this case, $s = \sqrt{u^2 + v^2}$ obeys Rayleigh probability density distribution [21] of the functional form

$$P_{\mathrm{R}}(s; s_0) = \frac{2}{s_0}\left(\frac{s}{s_0}\right)\exp\left[-\left(\frac{s}{s_0}\right)^2\right], \quad (1)$$

where the only free parameter is s_0 (the so called scale parameter).

A trivial consequence of the basic assumptions behind a Rayleigh distribution is that the mean vector wind should be zero as well. However, it is well known that the long range vectorial averages are significantly different from zero, in particular over the oceans (see e.g. [22,23]). Actually, these nonzero values define the prevailing wind systems (e.g. trade winds). This is the first reason why the Rayleigh distribution has a limited applicability, especially for sea winds [8,9].

The mean vector wind, however, is often close to zero over land [22,23]. Therefore, a next plausible test on the validity of the basic assumptions is to determine the normalized third and fourth central moments, the skewness (Sk) and kurtosis (K) for the individual wind components u and v:

$$\mathrm{Sk}(x) \approx \frac{\sum_{i=1}^{n}(x_i - \bar{x})^3}{n\sigma_x^3}, \quad K(x) \approx \frac{\sum_{i=1}^{n}(x_i - \bar{x})^4}{n\sigma_x^4} - 3, \quad (2)$$

where x_i is either u_i or v_i, and n is the number of observations. Eq. (2) are not exact equalities because they are biased estimators of skewness and kurtosis, but in this case, the sample size is very large ($n = 64240$) and Eq. (2) can be considered to be very good approximations. The results are shown in Fig. 2. The maps clearly illustrate that very few geographical locations exhibit a pure Gaussian probability distribution ($\mathrm{Sk} = 0, K = 0$) for the individual wind vector components.

The standard method to test interdependence of the components u and v is based on computing the correlation coefficient r_{uv} defined as

$$r_{uv} = \frac{\sum_{i=1}^{n}(u_i - \bar{u})(v_i - \bar{v})}{(n-1)\sigma_u\sigma_v}, \quad (3)$$

where an overline indicates average value, and σ_u and σ_v are the corresponding standard deviations, as before. Fig. 3 shows that the assumption of independence fails in general; strong correlations of magnitude 0.6–0.8 are present at several geographical locations. (Note that $r_{uv} \approx 0$ does not necessarily mean statistical independence.)

Nonzero correlations are usually taken into account by considering joint probability distributions. For example, when u and v are assumed to be Gaussian random variables with mean values $\bar{u}$ and $\bar{v}$, standard deviations σ_u and σ_v, and correlation coefficient r_{uv}, then the joint PDF can be written as

$$P(U, V) = \frac{1}{2\pi\sigma_u\sigma_v\sqrt{1 - r_{uv}^2}}\exp\left(-\frac{U^2 - 2r_{uv}UV + V^2}{2(1 - r_{uv}^2)}\right), \quad (4)$$

where $U = (u - \bar{u})/\sigma_u$ and $V = (v - \bar{v})/\sigma_v$ denote standardized variables.

There are two plausible methods to obtain standardized velocity components U and V. First, the mean values $\bar{u}$ and $\bar{v}$ can be computed over the whole length of the time series assuming a well defined prevailing wind. The second way is to consider velocity

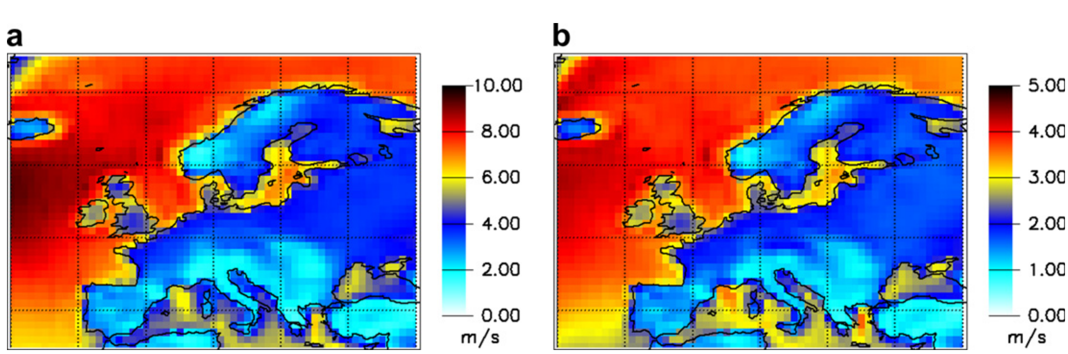

Fig. 1. (a) Average value, and (b) standard deviation of ERA-40 surface wind speeds in the period 1958–2002, in units of m/s. Note that the color scales are different by a factor of 2. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)